RICE UNIVERSITY
School of Engineering

# Databases
# COMP 630, Fall 2019

*Professor: John Greiner*
*Contact Information: greiner@rice.edu, Duncan Hall 3093*

*TA: Xiaochang (Linus) Huang*
*Contact Information: xh24@rice.edu*

## Course Materials:

All required course materials will be available via the Canvas web site.

## Software Requirements:

All the course software will be available via the web-based programming environment Vocareum. All programming-based assignments will be submitted via Vocareum. We will use the database systems PostgreSQL, MongoDB, and Neo4J, plus the tool Jupyter Notebook, plus libraries to link the database systems to Jupyter.

There will be one assignment that requires a small amount of programming in an application language (preferably Python or Java). We assume students are at least somewhat familiar with programming in such a language.

Downloading the course software onto your own computer is not necessary, but optional. All of the software is available for free.

## Course Overview:

The course has five learning objectives, i.e., high-level goals:

- (Big picture) Understand the trade-offs of relational and non-relational databases
- (Queries) Manage data and understand the costs of doing so
- (Design) Design complex databases and understand design trade-offs
- (Real-world data) Curate and merge data from real-world sources
- (Communication) Explain concepts and implementation and design decisions

## Course Requirements:

| Activity | Deadline | Grade weight | Time estimate |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Videos + in-video questions | | 0% | 2 hrs/wk |
| Quizzes | Wed. night | 10% | 1 hr/wk |
| Exercise sets | Wed. night | 17% | 1 hrs/wk |
| Feedback survey | Wed. night | 3% | 0.25 hrs/wk |
| Video summary + Q&A session | Thursday | 0% | 1 hr/wk |
| Assignments | Sun. night | 45% | 8 hrs each |
| Final exam | | 25% | 3 hrs |

The time required for each activity, especially the assignments, can vary tremendously by student. The times listed in the above table are only estimates.

Your activities in this class are organized around the following schedule. However, the material is released early so that you will have lots of flexibility to do the activities prior to the deadlines.

**Monday - Wednesday:** Each week will have two sets of material, or topics, each consisting of

- A set of video lectures -- Some videos contain in-video questions. These are not graded, but the course staff will see and review your answers.
- A "notes handout" -- This summarizes some of the important points from the videos, provides you a place to take notes, and repeats the information needed to complete the in-video questions.
- A quiz on the video content
- A set of exercises going into greater depth than the quiz

Each week will also have a short feedback survey. You will provide feedback to the instructor and, in doing so, also reflect on the week's material.

The quizzes, exercises, and feedback survey are all due Wednesday night.

**Thursday:** We will address any issues that students have had with the week's materials.

- We will host a live interactive session with group problem-solving and Q&A discussion. Participation is encouraged, but optional, and a video recording will be posted. This live session will be scheduled based upon a survey at the very beginning of the course.
- If there is a need, we'll precede this with a supplementary video on common issues.

**Sunday:** Roughly every two weeks, an assignment will be due. These will be due Sunday night.

**Other:** There will also be a final exam after the last week.

## Assignment Details:

| Assignment | Due Date | Points | Summary |
|---|---|---|---|
| 0 | 9/01/2019 | 10 | Summary of yourself |
| 1 | 9/15/2019 | 55 | Queries<br>Short answer problems on basic concepts |
| 2 | 9/25/2019 | 10 | Designing a genealogical DB: entity-relationship design and table schemas<br>Short answer problems on design & normalization |
| | 10/06/2019 | 90 | |
| 3 | 10/20/2019 | 100 | Complex SQL queries on the genealogical DB<br>Short answer problems on queries |
| 4 | 11/03/2019 | 100 | Parts of a transactional application on the genealogical DB<br>Short answer problems on DB/application integration & transactions |
| 5 | 11/17/2019 | 100 | Short answer problems on costs & efficiency |
| 6 | 12/08/2019 | 100 | OLAP, MongoDB, and Neo4J design of genealogical DB<br>MongoDB and Neo4J queries<br>Short answer problems |

## Course Outline:

The course will consist of six modules. Each module concludes with an assignment. The following summarizes the schedule and topics.

| Week | Module | Topics | Assignment due |
|---|---|---|---|
| 1 | 0 | Course overview | |
| | 1 | Database & relational concepts 1 | 0: 9/01/2019 |

| | | | |
|---|---|---|---|
| 2 | 1 | Database & relational concepts 2 | |
| | 1 | SQL 1, Simple queries | |
| 3 | 1 | SQL 2, Table and data operations | |
| | 2 | Data modeling overview | 1: 9/15/2019 |
| 4 | 2 | Entity-relationship modeling basics | |
| | 2 | Entity-relationship modeling classes | |
| 5 | 2 | Implementing ERDs in SQL | 2 initial milestone: 9/25/2019 |
| | 2 | Normalization background | |
| 6 | 2 | Normalization | |
| | 3 | SQL 3, Computation in queries | 2: 10/6/2019 |
| 7 | 3 | SQL 4, Multisets and quantifiers | |
| | 3 | SQL 5, More joins & binding forms | |
| 8 | 4 | SQL 6, Functions & triggers | |
| | 4 | Database/application interface | 3: 10/20/2019 |
| 9 | 4 | Transactions 1, Concurrency | |
| | 4 | Transactions 2, Recovery | |
| 10 | 4 | Security; Using real-world data | |
| | 5 | Data access costs 1, Storage | 4: 11/03/2019 |
| 11 | 5 | Data access costs 2, Indices | |
| | 5 | Query costs 1, Optimization | |
| 12 | 5 | Query costs 2, Tuning | |
| | 6 | Data warehouses, data lakes, & OLAP | 5: 11/17/2019 |
| 13 | 6 | NoSQL | |
| | 6 | MongoDB 1, Basics | |

| | | | |
|---|---|---|---|
| 14 | 6 | MongoDB 2, More | |
| | 6 | Neo4J | 6: 12/08/2019 |
| | | | Exam: 12/15/2019 |

## Honor Code:

All Masters of Computer Science students are sworn to adhere to the tenets set forth in Rice University's Honor Code. All written assignments for this class submitted to the instructor must carry the signed Honor Code pledge: "On my honor, I have neither given nor received any unauthorized aid on this (examination, quiz or paper)." (with signature and/or student ID number). Failure to include this pledge will result in an ungraded and returned assignment.

Please be aware of the Honor Code addendum that is part of the Student Handbook. This will provide more specific language about how the Honor Code should be adhered to as part of the Computer Science program.

In particular, for this course:

- Students may collaborate on solving non-exam problems, but must write their solutions independently.
- Students may consult printed and online sources when solving non-exam problems, but must cite such sources and may copy at most two lines of code for any problem.

## Attendance Policy:

Attendance is not required, but strongly encouraged for all live class sessions. The benefit of this program is the time spent during these live sessions interacting with your classmates and faculty, with much of the learning coming from these interactions. You will not be penalized for not attending, but please note that attendance in these live sessions is a major part of not only learning the content, but also the experience as a whole.

## Disability-based Accommodations:

If you have a documented disability that will impact your work in this class, please contact your instructor to discuss your needs. Additionally, you will need to register with the Disability Support Services (DSS) Office in the Allen Center. The phone number for DSS is (713) 348-5841.